MLSP 2012

2D Sound Source Localization on the Binaural Manifold

Antoine Deleforge* and Radu Horaud*

INRIA Grenoble Rhône-Alpes, France

Motivation

Traditional approaches to SSL

$$TDOA = \frac{||m_1 - s|| - ||m_2 - s||}{C} \propto \sin(\theta)$$

- Simplified model of sound propagation, based on the system geometry
- Mainly using the time difference of arrival (TDOA)

Pros:
- Simple and fast method
- Works well in anechoic, clean set-ups

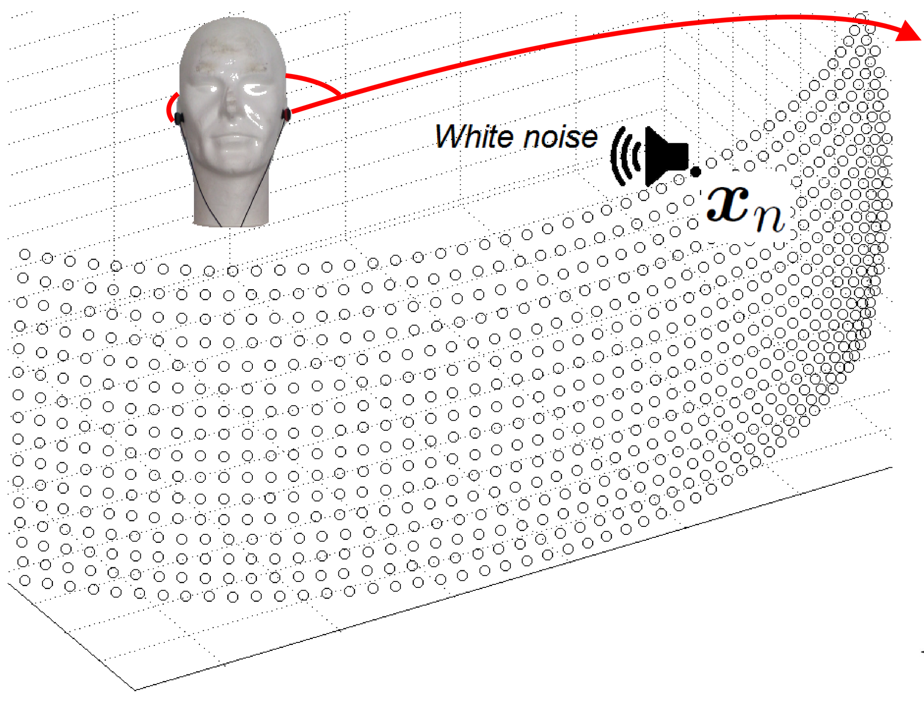
Cons:
- 1D localization only
- Poor accuracy due to approximations
- Need for external parameters
- Suffers in more complex set ups (ac. head)
- Performances depend on the emitted sound

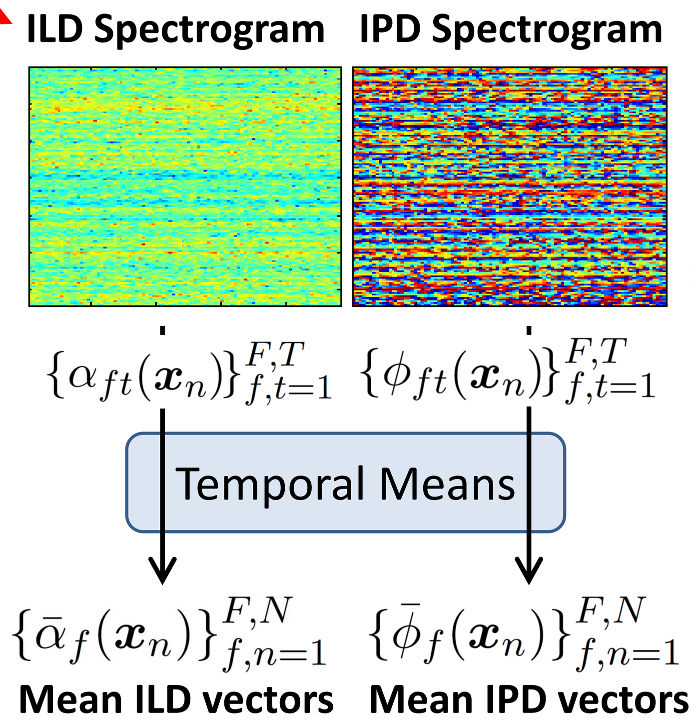
Challenges Addressed

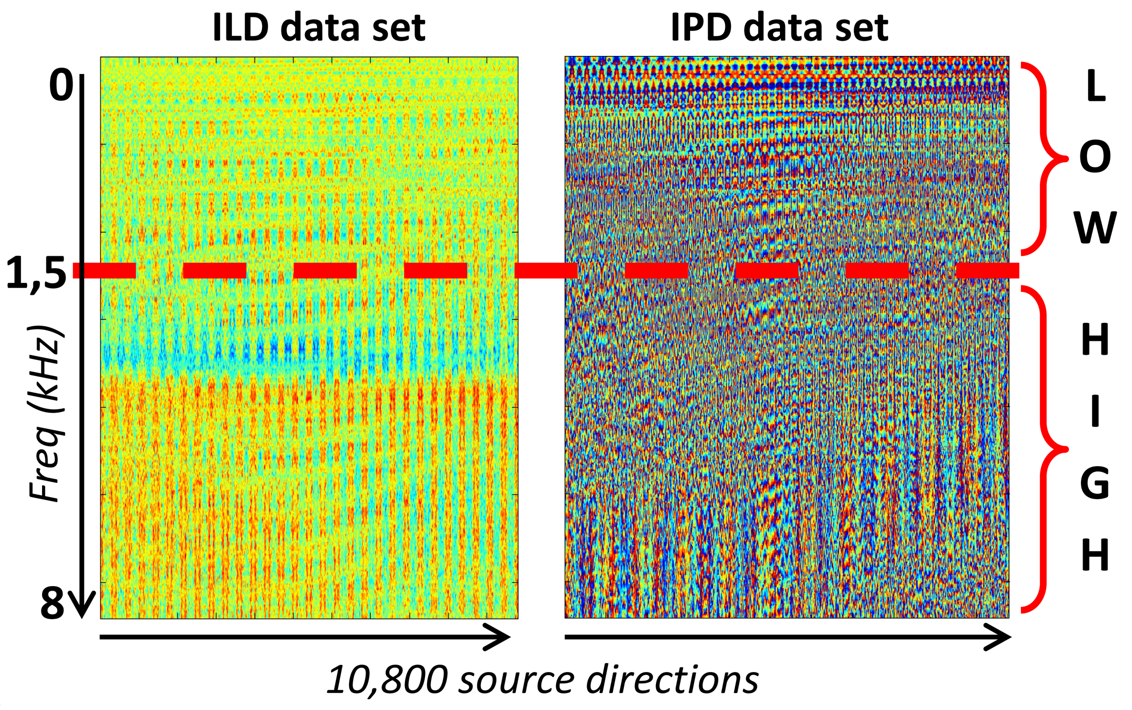
- Can we use spectral interaural cues for accurate 2D sound source localization?
- How to obtain a mapping function from cues to direction using a real world complex system with filtering effects?
- How to deal with the spectral sparsity of real world sounds such as speech?

An Audio-Motor Sound Database

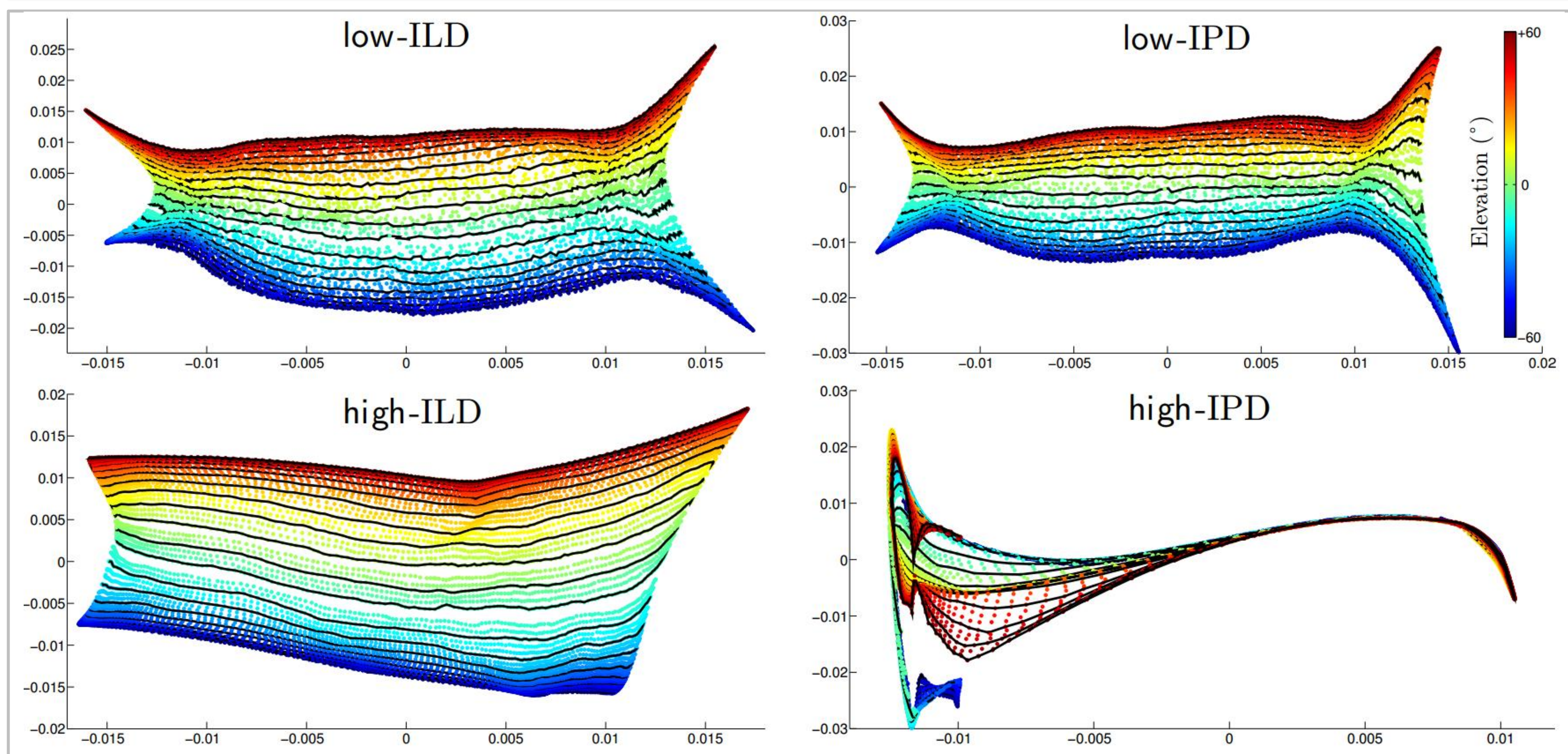
- 10,800 motor states
- 180 azimuth, 60 elevations, 2° steps
- White noise + rnd speech from TIMIT
- Data available online**



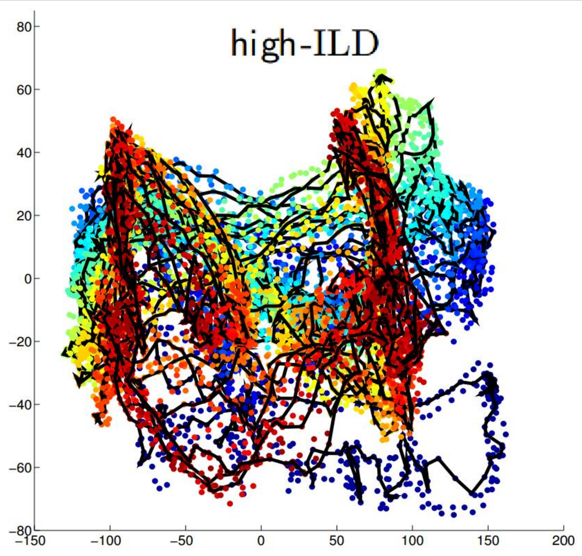




Intrinsic Manifold Structure of Spectral Interaural Data



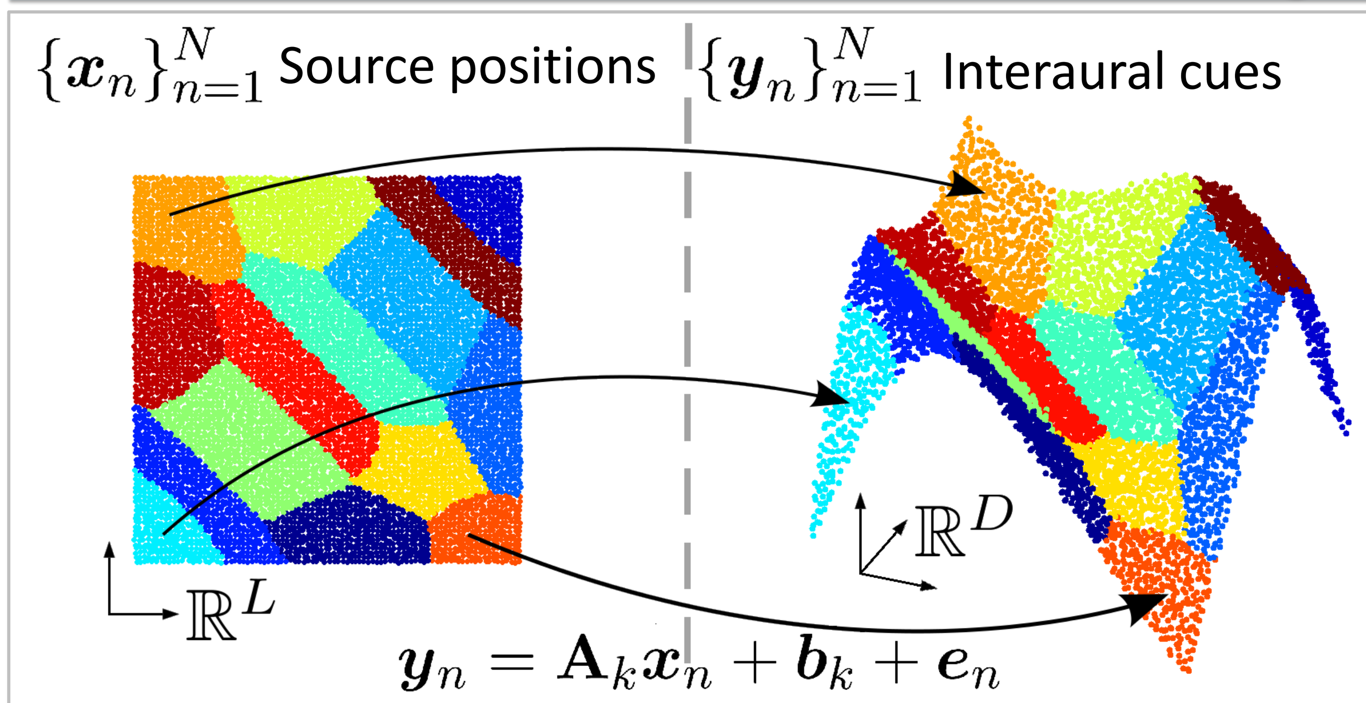
- Non linear dimensionality reduction on interaural data sets using the local tangent space alignment (LTSA) algorithm **(Zhang & Zha, 2003)**
- The smoothness of low-ILD, high-ILD and high-IPD representations shows the existence of an intrinsic manifold structure of spectral interaural data and of a homeorphism between 2D sound source direction and these cues
- Ambiguities in high-IPD cues confirming the Duplex theory

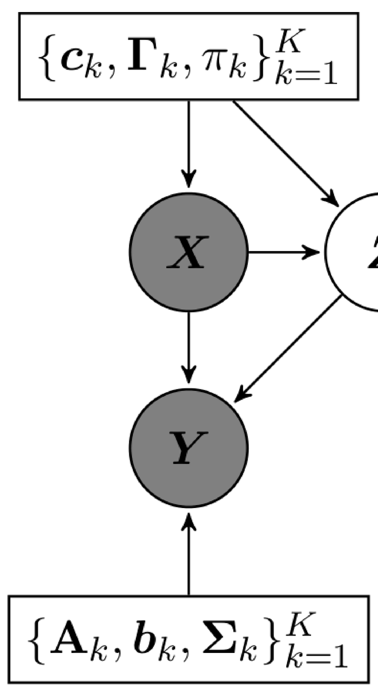


- Linear dimensionality reduction with principal component Analysis (PCA)
- Emphasizes the highly non-linear structure of interaural cues and the complexity of their relation to 2D sound source positions

Need for a locally linear mapping learning technique to perform accurate 2D sound source localization from spectral interaural cues

Probabilistic Piecewise Affine Regression (PPAR) - Learning





Generative Model

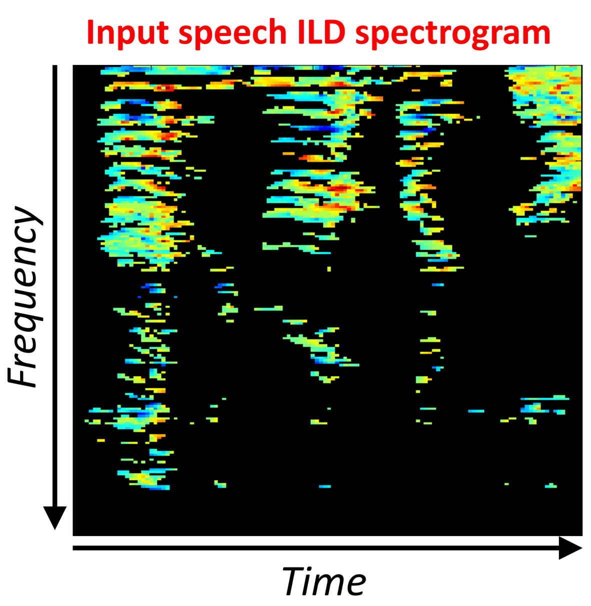
$$p(y_n | z_{kn} = 1, x_n; \Theta) = \mathcal{N}(y_n; A_k x_n + b_k, \Sigma)$$

$$p(z_{kn} = 1 | x_n; \Theta) = \frac{\pi_k \mathcal{N}(x_n; c_k, \Gamma_k)}{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_n; c_k, \Gamma_k)}$$

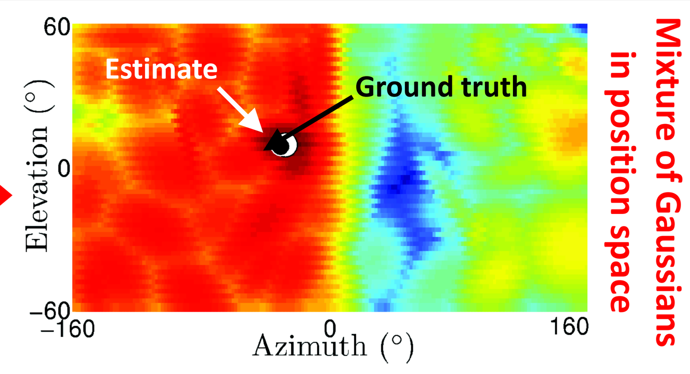
$$p(x_n; \Theta) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n; c_k, \Gamma_k)$$

Maximum likelihood estimation of Θ with a closed form EM algorithm

PPAR - Spectrogram Inversion



- 80% of missing data $\chi = \{\chi_{td}\}_{t,d=1}^{T,D}$
- Noisy and redundant observations



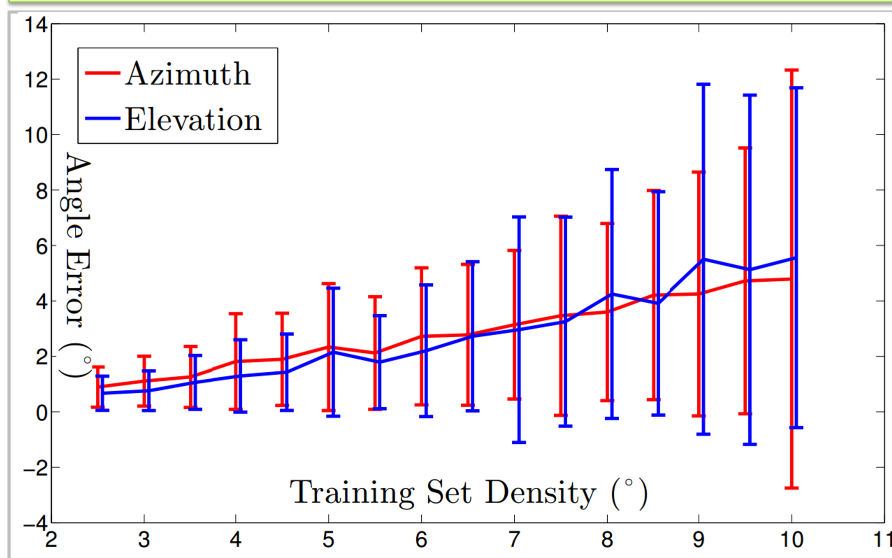
$p(x | Y, \chi; \widetilde{\Theta}) = \sum_{k=1}^{K} \alpha_k \mathcal{N}(x; \mu_k, V_k)$ with:

$$\mu_k = V_k \left(\widetilde{\Gamma}_k^{-1} \widetilde{c}_k + \sum_{d,t=1}^{D,T} \frac{\chi_{dt}}{\widetilde{\sigma}_d^2} \widetilde{a}_{dk} (y_{dt} - \widetilde{b}_{dk}) \right)$$

$$V_k = \left(\widetilde{\Gamma}_k^{-1} + \sum_{d,t=1}^{D,T} \frac{\chi_{dt}}{\widetilde{\sigma}_d^2} \widetilde{a}_{dk} \widetilde{a}_{dk}^{\top} \right)^{-1}$$

$$\alpha_k \propto (|V_k| / |\widetilde{\Gamma}_k|)^{\frac{1}{2}} \exp\Big(-\tfrac{1}{2}\Big(\sum_{d,t=1}^{D,T} \tfrac{\chi_{dt}}{\widetilde{\sigma}_d^2} (y_{dt} - \widetilde{b}_{dk})^2 + \widetilde{c}_k^{\top} \widetilde{\Gamma}_k^{-1} \widetilde{c}_k + \mu_k^{\top} V_k^{-1} \mu_k\Big)\Big)$$

Results



- Average and standard deviation of angular errors in white noise localization as a function of the training set's density.

Training density	Azimuth			Elevation		
	Avg	Std	Out	Avg	Std	Out
$\delta = 2°$	2.0°	1.8°	1.2%	1.3°	1.2°	1.2%
$\delta = 6°$	5.6°	5.2°	1.2%	4.5°	5.3°	1.4%

- Localization results with real world speech recordings

Conclusion

- A general framework to study and map together data presenting a non-linear manifold structure
- Insights on the richness of interaural spectral cues for 2D sound localization

Future work

- Localization of multiple sound sources using mixed inverse PPAR models
- Robustness to reverberations

*e-mails: {Antoine.Deleforge, Radu.Horaud}@inria.fr

**http://perception.inrialpes.fr/~Deleforge/CAMIL_Dataset/